```
#Regn no-1061
#Name-G.Vamshikrishna
#Date-12-10-2022
import numpy as np
import pandas as pd
```

```
#df=pd.read_csv("/content/Enrollments_28092022.csv")
from google.colab import files
files.upload()
```

Choose Files No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

```
Saving Enrollments_28092022.csv to Enrollments_28092022.csv
{'Enrollments_28092022.csv':
b'\xef\xbb\xbfStudentNo,DEGREE,INTERMEDIATE,SSC,INTERNSHIP\r\n1001,8.1,76,92,Dat
 Science\r\n1002,8.1,76,92,MEAN Stack Web Development\r\n1003,7.8,94.6,92,MEAN
Stack Web Development\r\n1004,9.03,89.5,89,Data Science\r\n1005,8.38,87,90,MEAN
Stack Web Development\r\n1006,8.9,83,64,Cloud Computing Services
(AWS)\r\n1007,7.2,87,85,Cloud Computing Services (AWS)\r\n1008,7.5,81,70,Cloud
Computing Services (AWS)\r\n1009,8.69,82.6,88,Data
Science\r\n1010,8.3,87,98,MEAN Stack Web Development\r\n1011,8,88,85,MEAN Stack
Web Development\r\n1012,8.4,87,90.3,Cloud Computing Services
(AWS)\r\n1013,7.66,92.1,85,MEAN Stack Web Development\r\n1014,7.24,80,85,Cloud
Computing Services (AWS)\r\n1015,9.07,95,98,Cloud Computing Services
(AWS)\r\n1016,8.1,84,97,Cloud Computing Services
(AWS)\r\n1017,9.53,98.2,93.1,MEAN Stack Web
Development\r\n1018,9.08,95.2,83,MEAN Stack Web
Development\r\n1019,8.8,95.6,78.85,Cloud Computing Services
(AWS)\r\n1020,9.16,98,87,Data Science\r\n1021,9,96.8,99,Data
Science\r\n1022,8,95.2,98,Data Science\r\n1023,7,84,83,Data
Science\r\n1024,8.78,95.7,95,Data Science\r\n1025,9.08,97.2,97,Data
Science\r\n1026,7.5,79,85,Data Science\r\n1027,8.4,94.9,93,Data
Science\r\n1028,8.6,75,85,Cloud Computing Services
(AWS)\r\n1029,8.92,89.7,83,Cloud Computing Services
(AWS)\r\n1030,8.65,91,90,Data Science\r\n1031,8.56,78.2,88.35,Cloud Computing
Services (AWS)\r\n1032,8.6,92,60,Data Science\r\n1033,8.7,91.4,93,Data
Science\r\n1034,8.02,86,87,MEAN Stack Web Development\r\n1035,8,88,88,Data
Science\r\n1036,8.62,92.7,88,Cloud Computing Services
(AWS)\r\n1037,7.5,79.2,75,Data Science\r\n1038,7.5,83,78,Data
Science\r\n1039,8.5,80,87,Data Science\r\n1040,9.03,96.5,97,Data
Science\r\n1041,9.34,97.7,97,Data Science\r\n1042,8.3,82,92,Data
Science\r\n1043,7.8,88,87,Data Science\r\n1044,7.8,88,87,Data
Science\r\n1045,9,96,86,Data Science\r\n1046,8.3,95,95,Data
Science\r\n1047,8.45,95,93,Data Science\r\n1048,7.5,73.6,75,Data
Science\r\n1049,7.9,93,87,Cloud Computing Services
(AWS)\r\n1050,8.72,91.7,75,Data Science\r\n1051,8.27,94,95,Data
Science\r\n1052,7.6,68,68,Data Science\r\n1053,7,86.3,92,MEAN Stack Web
Development\r\n1054,8.39,93.7,93.7,Cloud Computing Services
(AWS)\r\n1055,8.89,95.1,95,Data Science\r\n1056,8.66,95,80,Cloud Computing
Services (AWS)\r\n1057,8.68,94.3,83,Data Science\r\n1058,8.2,95,95,Data
Science\r\n1059,7.8,70,75,Data Science\r\n1060,8.02,95.5,93,Data
Science\r\n1061,8.8,97.3,92,Data Science\r\n1062,7.8,80.3,92,Data
Science\r\n1063,7.4,81.2,85,Data Science\r\n1064,8.19,91.6,93,Cloud Computing
Services (AWS)\r\n1065,8,91.5,88,Data Science\r\n1066,8.3,94,85,Data
Science\r\n1067,7,81.7,93,Data Science\r\n1068,8.45,95.2,90,Data
Science\r\n1069,8.3,93,88,Data Science\r\n1070,8,80,60,Data
Science\r\n1071,8.88,6.85,Data Science\r\n1072,8.09,90.2,82,Data
```

```
df=pd.read_csv("/content/Enrollments_28092022.csv")
df.head()
```

	StudentNo	DEGREE	INTERMEDIATE	SSC	INTERNSHIP
0	1001	8.10	76.0	92.0	Data Science
1	1002	8.10	76.0	92.0	MEAN Stack Web Development

```
df.tail()
```

	StudentNo	DEGREE	INTERMEDIATE	SSC	INTERNSHIP
292	2188	8.70	94.1	93.0	Data Science
293	2189	8.45	90.0	93.0	Data Science
294	2190	8.40	94.9	98.0	Data Science
295	2191	7.06	90.6	88.0	Cloud Computing Services (AWS)
296	2192	7.50	95.5	95.0	Cloud Computing Services (AWS)

Computing Services (AWS)\n\n2021 7.80 88 Cloud Computing Services

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 297 entries, 0 to 296
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   StudentNo     297 non-null    int64
 1   DEGREE        297 non-null    float64
 2   INTERMEDIATE  297 non-null    float64
 3   SSC           297 non-null    float64
 4   INTERNSHIP    297 non-null    object
dtypes: float64(3), int64(1), object(1)
memory usage: 11.7+ KB
```

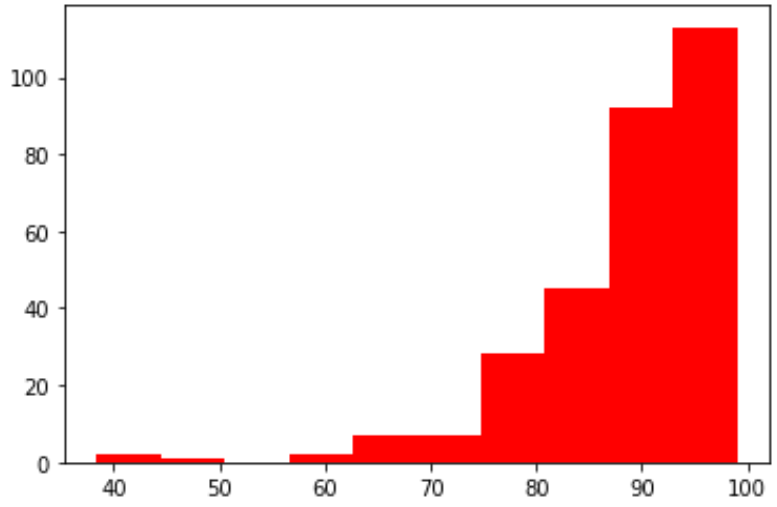
```
print("Number of rows: ",len(df))
print("Number of columns: ",len(df.axes[1]))
```

```
Number of rows:  297
Number of columns:  5
```

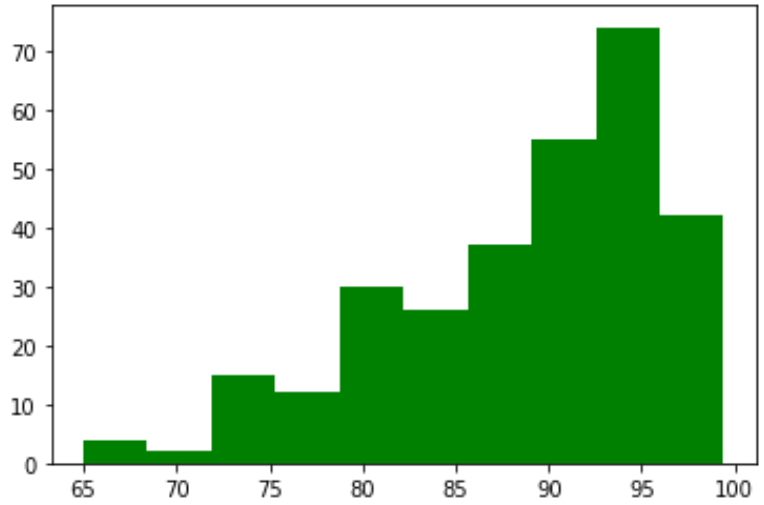
```
import matplotlib.pyplot as plt
```

```
plt.hist(df['DEGREE'])
plt.show()
```


```
plt.hist(df['SSC'],color='red')
plt.show()
```




```
plt.hist(df['INTERMEDIATE'],color='green')
plt.show()
```

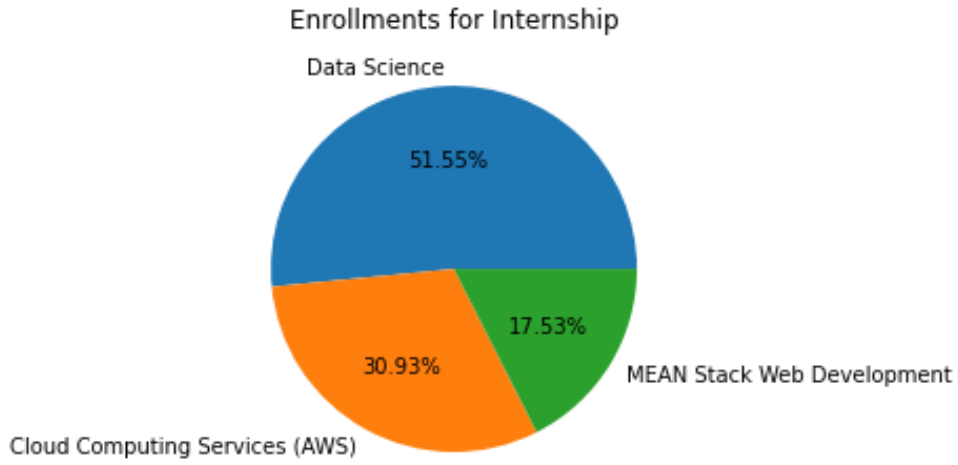



```
df['INTERNSHIP'].value_counts()
```

```
Data Science                      156
Cloud Computing Services (AWS)     90
MEAN Stack Web Development         51
Name: INTERNSHIP, dtype: int64
```

```
Internship=['Data Science','Cloud Computing Services (AWS)','MEAN Stack Web Developmer
```

```
Total=[150,90,51]
plt.pie(Total,labels=Internship,autopct='%1.2f%%')
plt.title("Enrollments for Internship")
plt.show()
```




```
print("SSC")
print("mean= ",df['SSC'].mean())
print("median= ",df['SSC'].median())
print("mode= ",df['SSC'].mode())
```

```
SSC
mean=  88.10673400673402
median=  90.0
mode=  0    95.0
dtype: float64
```

```
print("DEGREE")
print("mean= ",df['DEGREE'].mean())
print("median= ",df['DEGREE'].median())
print("mode= ",df['DEGREE'].mode())
```

```
DEGREE
mean=  7.928080808080809
median=  8.0
mode=  0    7.0
dtype: float64
```

```
print("INTERMEDIATE")
print("mean= ",df['INTERMEDIATE'].mean())
print("median= ",df['INTERMEDIATE'].median())
print("mode= ",df['INTERMEDIATE'].mode())
```

```
INTERMEDIATE
mean=  88.66262626262626
median=  90.8
mode=  0    95.0
dtype: float64
```

```
cv = lambda x: np.std(x, ddof=1) / np.mean(x) * 100
```

```
print("SSC")
print("min= ",df['SSC'].min())
print("max= ",df['SSC'].max())
print("Range= ",df['SSC'].max()-df['SSC'].min())
print("standard deviation= ",df['SSC'].std())
print("coefficent variation=",cv(df['SSC']))
```

```
SSC
min=  38.4
max=  99.0
Range=  60.6
standard deviation=  9.027984183574615
coefficent variation= 10.24664491920062
```

```
print('DEGREE')
print("min= ",df['DEGREE'].min())
print("max= ",df['DEGREE'].max())
print("Range= ",df['DEGREE'].max()-df['DEGREE'].min())
print("standard deviation= ",df['DEGREE'].std())
print("coefficent variation=",cv(df['DEGREE']))
```

```
DEGREE
min=  5.8
max=  9.53
Range=  3.7299999999999995
standard deviation=  0.7855786429497713
coefficent variation= 9.90881225818308
```

```
print('INTERMEDIATE')
print("min= ",df['INTERMEDIATE'].min())
print("max= ",df['INTERMEDIATE'].max())
print("Range= ",df['INTERMEDIATE'].max()-df['INTERMEDIATE'].min())
print("standard deviation= ",df['INTERMEDIATE'].std())
print("coefficent variation=",cv(df['INTERMEDIATE']))
```

```
INTERMEDIATE
min=  65.0
max=  99.4
Range=  34.400000000000006
standard deviation=  7.35573276879534
coefficent variation= 8.29631726338337
```

```
import scipy.stats as stats
```

```
print(stats.zscore(df['SSC']))
```

```
0      0.431972
1      0.431972
2      0.431972
3      0.099111
4      0.210065
          ...
292    0.542926
293    0.542926
294    1.097694
295   -0.011843
296    0.764833
Name: SSC, Length: 297, dtype: float64
```

```
print(stats.zscore(df['DEGREE']))
```

```
0      0.219213
1      0.219213
2     -0.163315
3      1.405052
4      0.576240
          ...
292    0.984271
293    0.665497
294    0.601742
295   -1.106886
296   -0.545844
Name: DEGREE, Length: 297, dtype: float64
```

```
print(stats.zscore(df['INTERMEDIATE']))
```

```
0     -1.724369
1     -1.724369
2      0.808539
3      0.114032
4     -0.226413
          ...
292    0.740450
293    0.182121
294    0.849392
295    0.263827
296    0.931099
Name: INTERMEDIATE, Length: 297, dtype: float64
```

```
#Inter quartile Range  for DEGREE
q3, q1 = np.percentile(df['DEGREE'], [75 ,25])
iqr = q3 - q1
iqr
```

```
1.1600000000000001
```

```
#Finding Inter-quartile Range for Intermediate
q3, q1 = np.percentile(df['INTERMEDIATE'], [75 ,25])
```

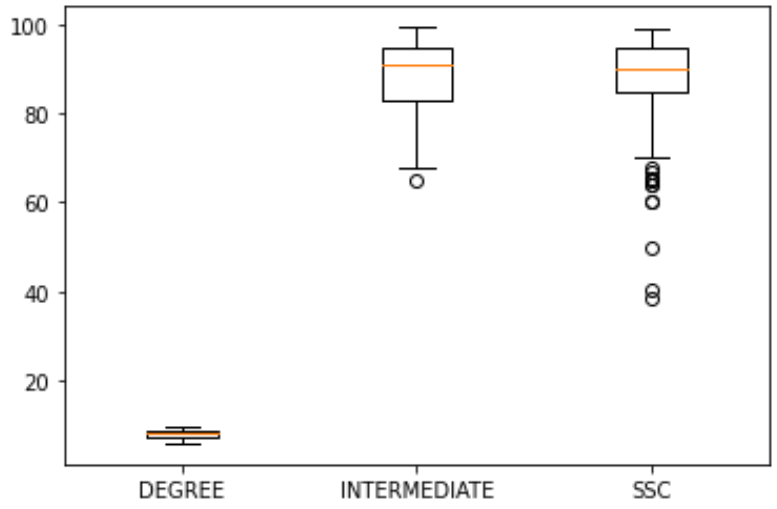
```
iqr = q3 - q1
iqr
```

```
11.599999999999994
```

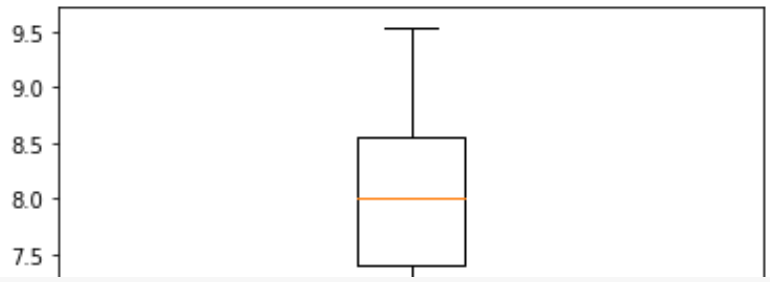
```
#Finding Inter-quartile Range for SSC
q3, q1 = np.percentile(df['SSC'], [75 ,25])
iqr = q3 - q1
iqr
```

```
10.0
```

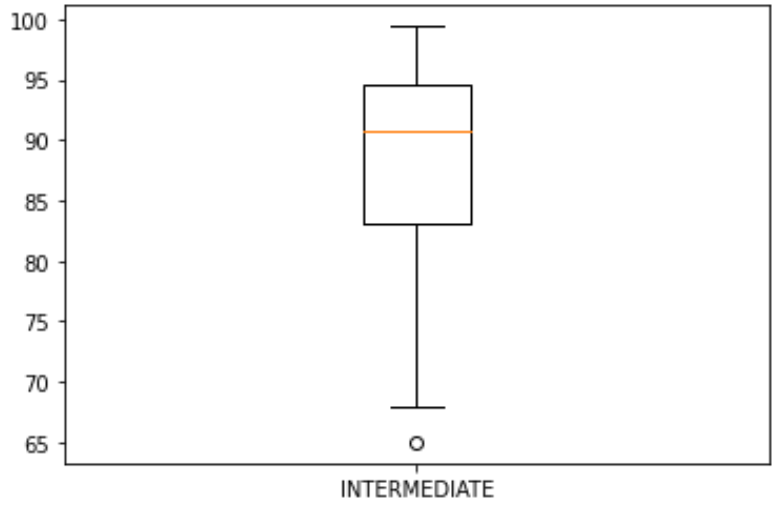
```
#BOXPLOT
import matplotlib.pyplot as plt
DEGREE=df['DEGREE']
INTERMEDIATE=df['INTERMEDIATE']
SSC=df['SSC']
columns=[DEGREE,INTERMEDIATE,SSC]
fig,ax=plt.subplots()
ax.boxplot(columns)
plt.xticks([1,2,3],["DEGREE","INTERMEDIATE","SSC"])
plt.show()
```



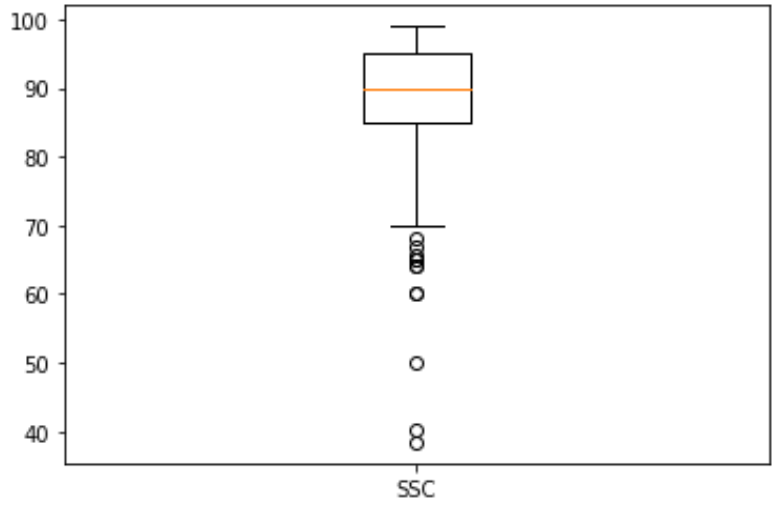

```
#BOxplot for DEGREE
DEGREE=df['DEGREE']
columns=[DEGREE]
fig,ax=plt.subplots()
ax.boxplot(columns)
plt.xticks([1],["DEGREE"])
plt.show()
```




```
#Boxplot for INTERMEDIATE
INTERMEDIATE=df['INTERMEDIATE']
columns=[INTERMEDIATE]
fig,ax=plt.subplots()
ax.boxplot(columns)
plt.xticks([1],["INTERMEDIATE"])
plt.show()
```




```
#BOXPLOT for SSC
SSC=df['SSC']
columns=[SSC]
fig,ax=plt.subplots()
ax.boxplot(columns)
plt.xticks([1],["SSC"])
plt.show()
```



```
#Finding Outliers
def outlier(a):
  q1=np.quantile(a,0.25)
  q3=np.quantile(a,0.75)
  med=np.median(a)
  iqr=q3-q1
  upper_bound=q3+(1.5*iqr)
  lower_bound=q1-(1.5*iqr)
  print(iqr,upper_bound,lower_bound)
  print('Inter-Quartile Range:',iqr)
  outliers=a[(a<=lower_bound)|(a>=upper_bound)]
  print('Outliers in the boxplot:\n{}'.format(outliers))
```

```
#Outlier for DEGREE
outlier(df['DEGREE'])
```

```
1.1600000000000001 10.3 5.66
Inter-Quartile Range: 1.1600000000000001
Outliers in the boxplot:
Series([], Name: DEGREE, dtype: float64)
```

```
#Outlier for INTERMEDIATE
outlier(df['INTERMEDIATE'])
```

```
11.599999999999994 111.99999999999999 65.60000000000001
Inter-Quartile Range: 11.599999999999994
Outliers in the boxplot:
271    65.0
Name: INTERMEDIATE, dtype: float64
```

```
#Outlier for SSC
outlier(df['SSC'])
```

```
10.0 110.0 70.0
Inter-Quartile Range: 10.0
Outliers in the boxplot:
5      64.0
7      70.0
31     60.0
51     68.0
69     60.0
82     65.6
86     50.0
107    64.0
236    38.4
237    67.0
243    40.2
270    65.0
288    65.0
Name: SSC, dtype: float64
```

```
#No.Of Students with 90% Percentile for DEGREE
Degree=pd.DataFrame(df[df['DEGREE']>=9.0])
len(Degree)
```

```
24
```

```
#No.Of Students with 90% Percentile for INTERMEDIATE
Intermediate=pd.DataFrame(df[df['INTERMEDIATE']>=90])
len(Intermediate)
```

```
164
```

```
#No.Of Students with 90% Percentile for SSC
Ssc=pd.DataFrame(df[df['SSC']>=90])
len(Ssc)
```

```
164
```

```

```